LEVERAGING LINKED DATA: SEMANTIC WEB TECHNOLOGIES APPLIED TO ACS DATA

Jonathan Ortiz
May 12, 2017
OUR PURPOSE

We’re building the most meaningful, collaborative, and abundant data resource in the world by dismantling the barriers between data and people.

Our platform helps people who work with data solve problems faster by creating new ways to discover, prep, and collaborate.
There are a **HUGE** number of **OPEN DATA SETS**
TOO MUCH OF DATA’S GROWTH IS HAPPENING IN SILOS
Finding it, understanding it, and using it can be a challenge.
This process happens **over and over and over again** as each data user does it individually.
So much HUMAN EFFORT is wasted on the WORKING & REWORKING of the SAME DATA
What is **LINKED DATA**?
It’s applying the **SAME** architecture as the **WWW** of linked documents to... **DATA**
First, break **DATA** into **ATOMIC FACTS**
(SUBJECT, PREDICATE, OBJECT)

(Turkey, "is a", Country)
(SUBJECT, PREDICATE, OBJECT)

(Turkey, "is a", Country)

(Ankara, "is a", City)
(SUBJECT, PREDICATE, OBJECT)

(Turkey, "is a", Country)

(Ankara, "is a", City)

(Ankara, "is the capital of", Turkey)
SUBJECT, PREDICATE, OBJECT
THE TRIPLE

City "is a" Ankara "is the capital of" Turkey "is a" Country

data.world
Refer to **ENTITIES** and **RELATIONSHIPS** via **URIs** so their **MEANINGS** can be discussed
SUBJECT, PREDICATE, OBJECT

THE CODE

```
<http://dbpedia.org/resource/Turkey>
  <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>

<http://dbpedia.org/resource/Ankara>
  <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
  <http://dbpedia.org/ontology/City> .

<http://dbpedia.org/resource/Ankara>
  <http://dbpedia.org/ontology/capital>
  <http://dbpedia.org/resource/Turkey> .
```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix dbpedia: <http://dbpedia.org/resource/> .
@prefix dbo: <http://dbpedia.org/ontology/> .

dbpedia:Turkey
  rdf:type dbo:Country .
dbpedia:Ankara
  rdf:type dbo:City ;
  dbo:capital dbpedia:Turkey .
PUTTING IT TOGETHER

City

"is a"

"is a"

"is the capital of"

Ankara

Turkey

Country

(dbpedia:Ankara rdf:type dbo:City)

(dbpedia:Ankara dbo:capital dbpedia:Turkey)

(dbpedia:Turkey rdf:type dbo:Country)
TURKEY vs TURKEY

(dbpedia:Turkey, rdf:type, dbo:Country)
(dbpedia:Turkey_(bird), rdf:type, dbo:Bird)
(dbpedia:Turkey, foaf:name, "Turkey")
(dbpedia:Turkey_(bird), foaf:name, "Turkey")
FIVE-STAR OPEN DATA

- PDF
- XLS
- CSV
- RDF
- LOD
WHY IT MATTERS

• Usability and availability
• Sharing knowledge with data
• Better starting point
Why Semantics & Data Linking is Vital to Artificial Intelligence

Artificial intelligence and cognitive computing seem to be in the news almost everyday – everything from autonomous vehicles to chatbots. But what’s really driving this shift in technology? It’s the rise of artificial intelligence and cognitive computing that is enabling new approaches to problem-solving and decision-making. It’s also enabling new ways of thinking about data – how it is collected, analyzed, and used to make decisions.

A key component of artificial intelligence and cognitive computing is the ability to understand the meaning of data. This is where semantics come in. Semantics refers to the study of meaning in language. By understanding the meaning of data, we can better understand the context in which it was generated and how it relates to other data.

In the world of artificial intelligence and cognitive computing, data is often collected from a variety of sources, each of which may have a different meaning to the data. By understanding the semantics of the data, we can better understand how it relates to other data and make more informed decisions.

Another important aspect of artificial intelligence and cognitive computing is the ability to link data from different sources. This is where data linking comes in. Data linking refers to the process of connecting data from different sources in order to create a more complete picture of the data.

By linking data, we can better understand how different sources of data are related to each other. This is important because it allows us to create a more complete picture of the world and make more informed decisions.

In conclusion, the rise of artificial intelligence and cognitive computing is being driven by the need to understand the meaning of data and how it relates to other data. This is where semantics and data linking come in. By understanding the meaning of data and linking it from different sources, we can make more informed decisions and better understand the world around us.
DEVELOP AN RDF SCHEMA FOR THE ACS PUMS
TYPICAL PUMS DATA RELEASES ARE TABULAR
### Semantic Modeling Provides Benefits

<table>
<thead>
<tr>
<th>Original</th>
<th>Enhanced</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flat, tabular data</td>
<td>URIs denote linkable concepts</td>
</tr>
<tr>
<td>Complex data dictionary</td>
<td>Metadata built into the data itself</td>
</tr>
<tr>
<td>Unwieldy raw files</td>
<td>Secure, queryable database</td>
</tr>
</tbody>
</table>
To make DISCOVERY of the data easier
To make the data INTEROPERABLE
To help the machines learn FASTER
The full **LINKED ACS** is available as an **AWS PUBLIC DATASET** thanks to our partners.
QUESTIONS
CONTACT INFO

data.world: @databeats
Twitter: @tweetsbyortiz
LinkedIn: linkedin.com/in/jonathanortiz
email: jonathan.ortiz@data.world
THANK YOU
APPENDIX
“AAA” Principal

ANYONE
Can say
ANYTHING
About
ANY TOPIC
2014 ACS PUMS DATA DICTIONARY
October 27, 2015

HOUSING RECORD

RT 1
Record Type
H. Housing Record or Group Quarters Unit

SERIALNO 7
Housing unit/GQ person serial number
0000001..9999999 .Unique identifier

DIVISION 1
Division code
0 . Puerto Rico
1 . New England (Northeast region)
2 . Middle Atlantic (Northeast region)
3 . East North Central (Midwest region)
4 . West North Central (Midwest region)
5 . South Atlantic (South region)
6 . East South Central (South region)
7 . West South Central (South Region)
8 . Mountain (West region)
9 . Pacific (West region)
NESTED PROPERTIES

RACIP 1
Recoded detailed race code
1 .White alone
2 .Black or African American alone
3 .American Indian alone
4 .Alaska Native alone
5 .American Indian and Alaska Native tribes specified; or American Indian or Alaska Native, not specified and no other races
6 .Asian alone
7 .Native Hawaiian and Other Pacific Islander alone
8 .Some Other Race alone
9 .Two or More Races

RAC2P 2
Recoded detailed race code
01 .White alone
02 .Black or African American alone
03 .Apache alone
04 .Blackfeet alone
05 .Cherokee alone
06 .Cheyenne alone
07 .Chickasaw alone
08 .Chippewa alone
09 .Choctaw alone
10 .Comanche alone
11 .Creek alone
12 .Crow alone
13 .Hopi alone
14 .Iroquois alone
15 .Lumbee alone
16 .Mexican American Indian alone
17 .Navajo alone
18 .Pima alone
...
68 .Two or More Races
COMBINED PROPERTIES

ESP 1
Employment status of parents
b. N/A (not own child of householder, and not child in subfamily)
   Living with two parents:
   1. Both parents in labor force
   2. Father only in labor force
   3. Mother only in labor force
   4. Neither parent in labor force
      Living with one parent:
      Living with father:
      5. Father in the labor force
      6. Father not in labor force
      Living with mother:
      7. Mother in the labor force
      8. Mother not in labor force
COMBINED PROPERTIES

ESP 1
Employment status of parents
b. N/A (not own child of householder, and not child in subfamily)
   Living with two parents:
     1. Both parents in labor force
     2. Father only in labor force
     3. Mother only in labor force
     4. Neither parent in labor force
   Living with one parent:
     Living with father:
       5. Father in the labor force
       6. Father not in labor force
     Living with mother:
       7. Mother in the labor force
       8. Mother not in labor force
ELEP 3
Electricity (monthly cost)
001. Included in rent or in condo fee
002. No charge or electricity not used
003. .999 $.3 to $999 (Rounded and top-coded)
### Mixed Properties

**ELEP 3**
- **Electricity (monthly cost)**
  - bbb: N/A (GQ/vacant)
  - 001: Included in rent or in condo fee
  - 002: No charge or electricity not used
  - 003..999: $3 to $999 (Rounded and top-coded)

**SEMP 6**
- **Self-employment income past 12 months (signed)**
  - bbbbb: N/A (less than 15 years old)
  - 00000: None
  - -10000..-00001: Loss of $1 to $10000 (Rounded and bottom-coded)
  - 000001: $1 or break even
  - 000002..999999: $2 to $999999 (Rounded and top-coded)
MIXED PROPERTIES

ELEP 3
Electricity (monthly cost)
bbb .N/A (GQ/vacant)
001 Included in rent or in condo fee
002 No charge or electricity not used
003..999 $3 to $999 (Rounded and top-coded)

SEMP 6
Self-employment income past 12 months (signed)
bbbbbb .N/A (less than 15 years old)
000000 .None
-10000...0001 .Loss of $1 to $10000 (Rounded and bottom-coded)
000001 .$1 or break even
000002..999999 .$2 to $999999 (Rounded and top-coded)

CONP 4
Condo fee (monthly amount)
bbbb .N/A (GQ/vacant/not owned or being bought)
0000 .Not condo
0001..9999 .$1 - $9999 (Rounded and top-coded)
MIXED PROPERTIES

ELEP 3
Electricity (monthly cost)
bbb .N/A (GQ/vacant)
001 .Included in rent or in condo fee
002 .No charge or electricity not used
003..999 .$3 to $999 (Rounded and top-coded)

SEMP 6
Self-employment income past 12 months (signed)
bbbbbb .N/A (less than 15 years old)
000000 .None
-10000..-00001 .Loss of $1 to $10000 (Rounded and bottom-coded)
000001 .$1 or break even
000002..999999 .$2 to $999999 (Rounded and top-coded)

CONP 4
Condo fee (monthly amount)
bbbb .N/A (GQ/vacant/not owned or being bought)
0000 .Not condo
0001..9999 .$1 – $9999 (Rounded and top-coded)

POVPIP 3
Income-to-poverty ratio recode
bbb .N/A
000..500 .Below 501 percent
501 .501 percent or more
MIXED PROPERTIES

ELEP 3
Electricity (monthly cost)
bobo .N/A (GQ/vacant)
001 .Included in rent or in condo fee
002 .No charge or electricity not used
003..999 .$3 to $999 (Rounded and top-coded)

SEMP 6
Self-employment income past 12 months (signed)
bobobob .N/A (less than 15 years old)
000000 .None
-1000...-00001 .Loss of $1 to $10000 (Rounded and bottom-coded)
000001 .$1 or break even
000002..999999 .$2 to $999999 (Rounded and top-coded)

CONP 4
Condo fee (monthly amount)
bobo .N/A (GQ/vacant/not owned or being bought)
0000 .Not condo
0001..9999 .$1 - $9999 (Rounded and top-coded)

POVIP 3
Income-to-poverty ratio recode
bob N/A
000..500 .Below 501 percent
501 .501 percent or more

FULP 4
Fuel cost (yearly cost for fuels other than gas and electricity)
bobo .N/A (GQ/vacant)
0001 .Included in rent or in condo fee
0002 .No charge or these fuels not used
0003..9999 .$3 to $9999 (Rounded and top-coded)
DATA INACCURACIES

ST 2
State Code
01. Alabama/AL
02. Alaska/AK
04. Arizona/AR
05. Arkansas/AR
06. California/CA
08. Colorado/CO
09. Connecticut/CT
10. Delaware/DE
11. District of Columbia/DC
12. Florida/FL
13. Georgia/GA
15. Hawaii/HI
16. Idaho/ID
17. Illinois/IL
18. Indiana/IN
19. Iowa/IA
20. Kansas/KS
21. Kentucky/KY
22. Louisiana/LA
23. Maine/ME
24. Maryland/MD
25. Massachusetts/MA
26. Michigan/MI
27. Minnesota/MN
28. Mississippi/MS
29. Missouri/MO
30. Montana/MT
31. Nebraska/NE
32. Nevada/NV
33. New Hampshire/NH
34. New Jersey/NJ
35. New Mexico/NM
36. New York/NY
37. North Carolina/NC
38. North Dakota/ND
39. Ohio/OH
40. Oklahoma/OK
41. Oregon/OR
42. Pennsylvania/PA
44. Rhode Island/RI
45. South Carolina/SC
46. South Dakota/SD
47. Tennessee/TN
48. Texas/TX
49. Utah/UT
50. Vermont/VT
51. Virginia/VA
53. Washington/WA
54. West Virginia/WV
55. Wisconsin/WI
56. Wyoming/WY
72. Puerto Rico/PR